



Enabling Grids for E-scienceE

ARDA and the SC4...

Ideas for discussion

Massimo Lamanna

preGDB meeting
CERN, 6th of September 2005



<http://arda.cern.ch>










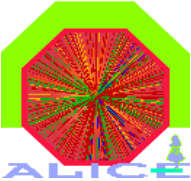

ARDA contribution to the SC4/preGDB workshop



- This talk is a summary of the “internal” discussions on the role and the interests of ARDA team in SC4
 - To stimulate discussion, understanding...
 - It is clearly work-in-progress
- Document available (this presentation follows it very closely):
http://lcg.web.cern.ch/LCG/activities/arda/public_docs/2005/Q3/SC4.doc (Aug 17)

ARDA prototypes

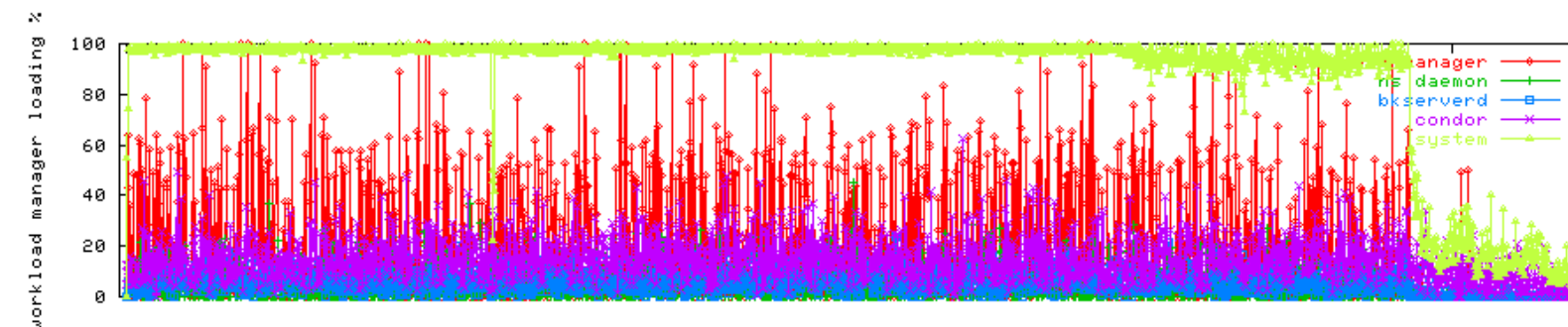
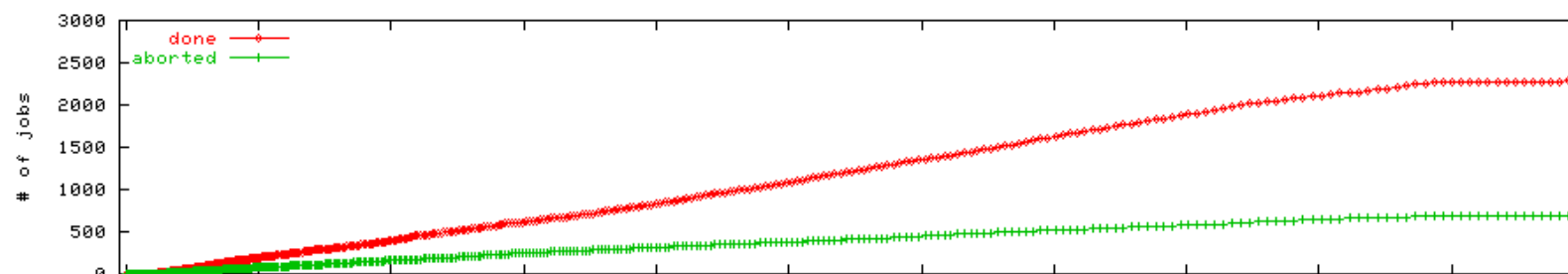
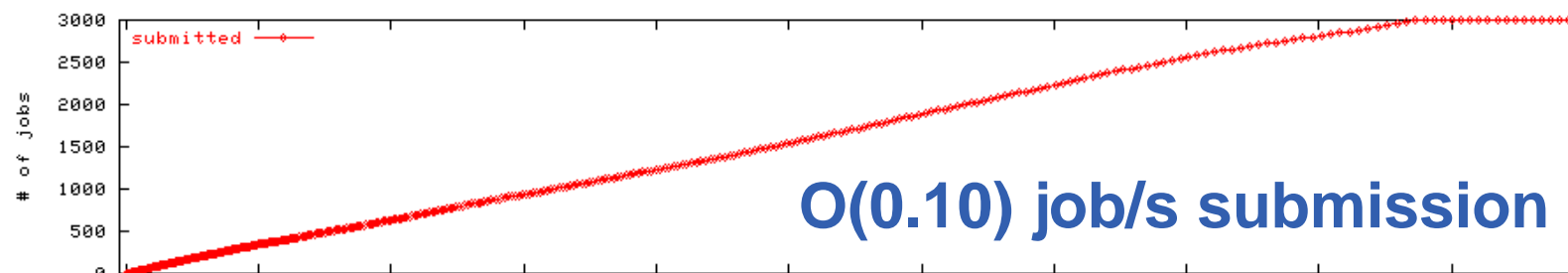


<i>LHC Experiment framework</i>	<i>Experiment- specific components</i>	<i>Middleware</i>
		
		
		
		

Scenarios 1,2,3: Submission performance



- On the current infrastructure job submission is limited to $o(10)$ jobs per minute.
 - If the system is frequently interrogated, this rate is goes down. While this submission rate is a tractable problem for the production, it is a heavy burden for user analysis.
 - Users do not expect to wait at least 10 minutes to submit 100 jobs.
 - 10^6 jobs a day is a realistic target (many numbers being discussed)
 - In the scenario where many individual users submit relative large bunches of jobs distribution will be even worse. Multiple client tools will also aggravate the problem...
 - First implementation of bulk submission system is now available and **tested (performance) by ARDA.**
 - Asynchronous submission (e.g. **CMS prototype**) is a necessity:
 - “Submit and go”
 - “MyFriends” service (used/using by CRAB/BOSS)
 - (Re)submission of job according general and experiment-specific policies
 - Implement experiment-specific policies

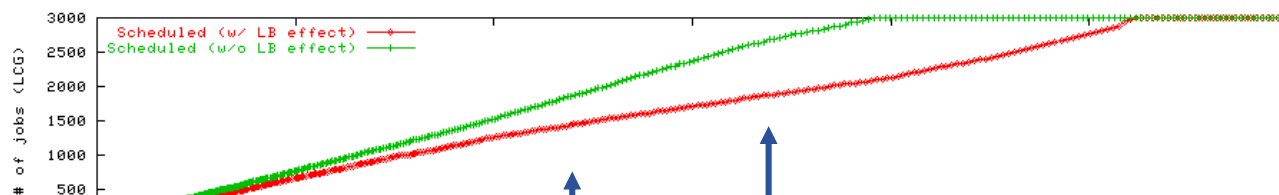


H-C Lee et al. reports:

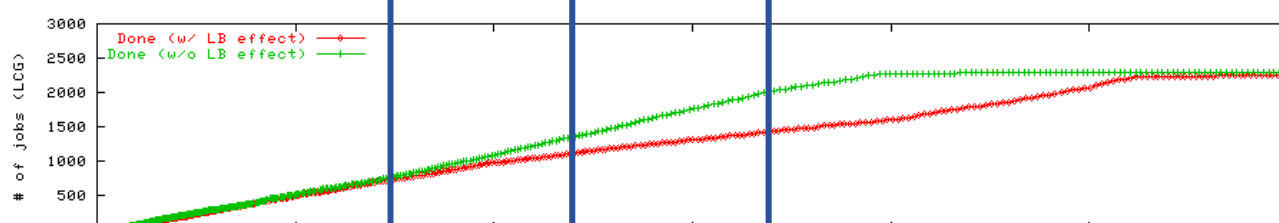
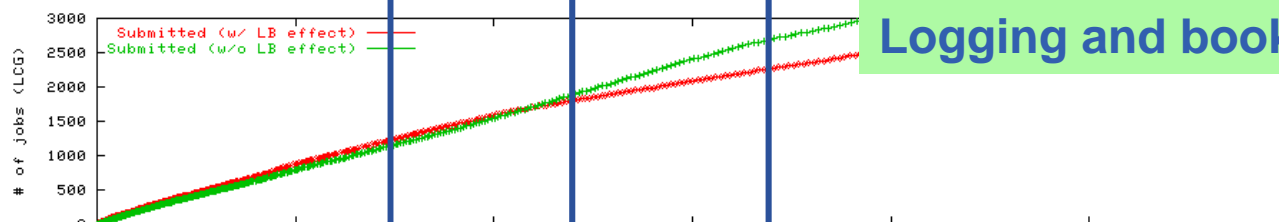
http://lcg.web.cern.ch/LCG/activities/arda/public_docs/2005/Q3/WMS%20Performance%20Test%20Plan.doc

http://lcg.web.cern.ch/LCG/activities/arda/public_docs/2005/Q3/perfWMS_rpt_2.ppt

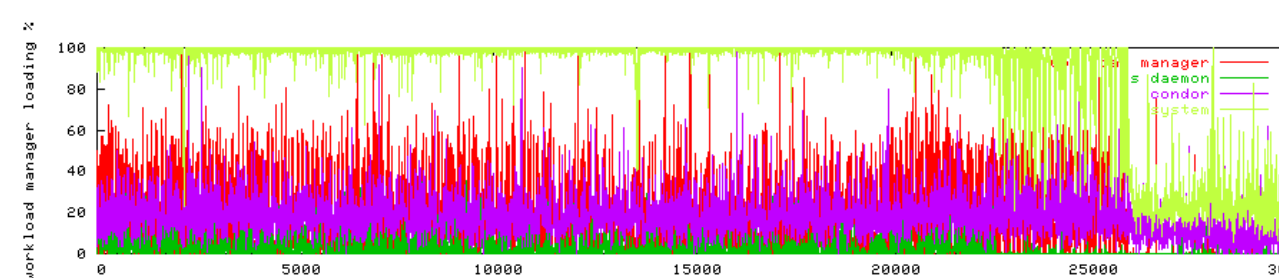
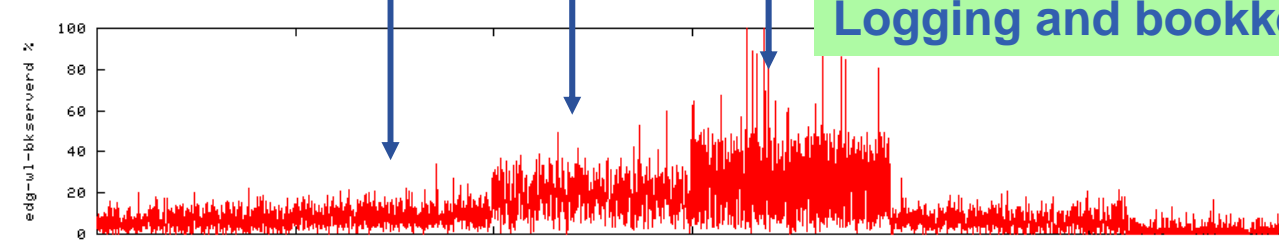
→ Bulk submissions preliminary: at least 3 times faster (pre-release)



Speed modulation induced by increasing Logging and bookkeeping load



Logging and bookkeeping (additional) load





- **How to mix long production and analysis?**
 - Maximise CPU delivery over ΔT ($\Delta T \sim o(10^5)$ or more)
 - Long jobs
 - Reduce latency
 - Queue behind production jobs? Preemption techniques?
 - Dedicated resources?
 - Pilot activity with ASCC (ATLAS) to get more experience



Scenario 4: I/O throughput within individual sites

- Analysis is often connected with jobs that require little CPU but lots of IO. In many cases the local IO throughput between the SE and the worker nodes at computing center will be the limitation.
- It is proposed to measure the throughput in a systematic way on all grid sites.
 - Since it is expected that the limiting factor for an effective analysis will be the bandwidth from each SE to the corresponding worker nodes, it is essential to characterize the different system and participate into the process of optimizing this part of the global service
- Could also be seen as a non-grid problem (analysis facility). Anyway it is a central problem to provide fast-turnaround pseudo interactive analysis, complementary to “batch” use.
 - Interest (in some cases part of the analysis system) at least in **ALICE**, **ATLAS** and **CMS** prototypes/activities

Scenario 5: Users' requests to FTS



- **The FTS service is the key of SC3.**
 - It is currently used by production manages of experiments. The effort is concentrated in distributing data to the lower Tier centers.
- **The typical analysis scenario would be the transfer of data to higher Tier centers for user analysis by users (groups of users) on demand.**
 - ARDA would like to experiment in having users triggering transfers to transfer sensible chunks of data (1-10 TB) compatibly with the experiments strategies and policies.

“Download in your favourite Tier2 a collections of data to be used by an analysis group...”

Scenario 6: Returning analysis results to the user



- Analysis jobs will typically return one or several small files to a SE close to the user.
- It has to be understood if this feature can be implemented by the FTS. Again this would require a "short" queue for the FTS which allows bypassing the "production" transfer.
- We would like to study, together with the operation people, a possible deployment scenario to provide this kind of service in an efficient way. The latency and reliability of such a system should be studied in several load scenarios.

Scenario 7: Analysis with “non official” software distribution



- **Today most of the analysis activities are based on software installation performed by VO software managers. This is a ridged schema and requires central coordination.**
 - Analysis would profit from a user driven installation. As an example users might like to perform their analysis on some specific release, that might be too old/too new or in any other way not supported by the central team.
 - More importantly, the latency introduced by the process of certification, packaging and distribution of the software prevents the efficient use of grid resources for final users (needing an essential new feature).
 - A final important aspect is the software installation on opportunistic resources, that might not even be known to the central installation team.
 - ARDA and all experiments have experience in different solution and ARDA would like to better investigate the existing mechanism and expose them to a significant users community.

Scenario 8: Use of the VO Box (Edge services) for Analysis



- **The mechanism of the VO Box (aka Edge services) has been proposed in the context of the LCG Baseline services working group.**
 - There is the expectation that some of the requirements implicit in the previous scenario would be satisfied by the use of the VO Box
 - ARDA would verify this assumption deploying and using the above mentioned analysis services.
 - What are the limits of the systems to deployed?
 - Control daemons?
 - Persistent services? (Data bases installed together with the service)